

UNITED STATES PATENT APPLICATION

NORMALIZING RECORDS

FIELD

5 This invention generally relates to information retrieval and more specifically relates to preparing a report from records based on a model norm.

BACKGROUND

10 The development of the EDVAC computer system of 1948 is often cited as the beginning of the computer era. Since that time, computer systems have evolved into extremely sophisticated devices, and computer systems may be found in many different settings. Computer systems typically include a combination of hardware, such as
15 semiconductors and circuit boards, and software, also known as computer programs. As advances in semiconductor processing and computer architecture push the performance of the computer hardware higher, more sophisticated and complex computer software has evolved to take advantage of the higher performance of the hardware, resulting in computer systems today that are much more powerful than just a few years ago.

20 One way that computers have become more complex and powerful has been by the connection of computers via networks. By connecting to a network, one computer, often called a client, may access information on various other computers, often called servers. The servers may provide information in a variety of formats, which causes the client difficulty in interpreting information. For example, one supplier of telephone
25 information may supply records including a name field, a telephone number field, and an email address field. Another supplier of telephone information may supply records including a name field, and street address field, and a telephone number field. Thus, although both suppliers provide telephone information, they provide records with some common fields, some different fields, and the fields may be ordered differently within the
30 records. This inconsistent nature of the different records causes the client difficulty in analyzing and interpreting the retrieved information.

Without a better way to organize information retrieved from different information suppliers, users will not be able to take advantage of the full power of computers.

SUMMARY

5 A method, apparatus, system, and signal-bearing medium are provided that in an embodiment receive records from information suppliers and prepare a report, which includes a subset of the characteristics of the records. The characteristics are selected based on a model norm, and the characteristics are ordered within the report based on a relative significance specified in the model norm. Also, the records are ordered within
10 the report based on a sort rule in the model norm. In this way, the records received from disparate information suppliers are normalized by the model norm to use a consistent format, which is more convenient for the requestor of the report.

BRIEF DESCRIPTION OF THE DRAWINGS

15 Fig. 1 depicts a block diagram of an example system for implementing an embodiment of the invention.

Fig. 2 depicts a block diagram of an example data structure for a supplier registry, according to an embodiment of the invention.

20 Fig. 3 depicts a block diagram of an example data structure for a data dictionary, according to an embodiment of the invention.

Fig. 4 depicts a block diagram of an example data structure for a favored norm, according to an embodiment of the invention.

Fig. 5 depicts a block diagram of an example data structure for a model norm, according to an embodiment of the invention.

25 Fig. 6A depicts a block diagram of an example data structure for records prior to being normalized, according to an embodiment of the invention.

Fig. 6B depicts a block diagram of an example data structure for a report, according to an embodiment of the invention.

Fig. 7 depicts a flowchart of example processing, according to an embodiment of the invention.

5

DETAILED DESCRIPTION

In an embodiment, a data assembler attracts clients who wish to rely on the data assembler's expertise in an area of interest. The data assembler generates a model norm, which reflects the data assembler's opinion on the format and organization of a report in the area of interest. When a client requests a report on the area of interest, the data assembler finds the appropriate information suppliers, retrieves records from the information suppliers, and a data normalizer creates a report in the format specified by the model norm. While the records may contain many fields or characteristics, the data normalizer selects those characteristics that are specified by the model norm and includes them in the report. The data normalizer further orders the selected characteristics as specified by the model norm and also orders the records within the report based on the model norm. In this way, records received from disparate information suppliers are organized into a consistent format, which is more convenient for the client.

Referring to the Drawing, wherein like numbers denote like parts throughout the several views, Fig. 1 depicts a block diagram of an example system 100 for implementing an embodiment of the invention. The system 100 includes an electronic device 102, an information supplier 104, and a client 106, all connected via a network 108. Although only one electronic device 102, one information supplier 104, one client 106, and one network 108 are shown, in other embodiments any number or combination of them may be present. Although the information supplier 104 and the client 106 are illustrated as being connected to the electronic device 102 via the same network 108, in other embodiments they may be connected to the electronic device 102 via different networks. Although the electronic device 102, the information supplier 104, the client 106, and the

network 108 are illustrated in Fig. 1 as being discrete, separate components, in other embodiments some or all of their functions and elements may be combined.

In an embodiment, the electronic device 102 functions as a server. The electronic device 102 includes a processor 110, a storage device 115, an input device 120, and an output device 122, all connected directly or indirectly via a bus 125. The processor 110 represents a central processing unit of any type of architecture, such as a CISC (Complex Instruction Set Computing), RISC (Reduced Instruction Set Computing), VLIW (Very Long Instruction Word), or a hybrid architecture, although any appropriate processor may be used. The processor 110 executes instructions and includes that portion of the electronic device 102 that controls the operation of the entire electronic device. Although not depicted in Fig. 1, the processor 110 typically includes a control unit that organizes data and program storage in memory and transfers data and other information between the various parts of the electronic device 102. The processor 110 reads and/or writes code and data to/from the network 108, the storage device 115, the input device 120, and/or the output device 122.

Although the electronic device 102 is drawn to contain only a single processor 110 and a single bus 125, embodiments of the present invention apply equally to electronic devices that may have multiple processors and multiple buses with some or all performing different functions in different ways.

The storage device 115 represents one or more mechanisms for storing data. For example, the storage device 115 may include read only memory (ROM), random access memory (RAM), magnetic disk storage media, optical storage media, flash memory devices, and/or other machine-readable media. In other embodiments, any appropriate type of storage device may be used. Although only one storage device 115 is shown, multiple storage devices and multiple types of storage devices may be present. Although the storage device 115 is shown in Fig. 1 as a single monolithic entity, the storage device 115 may in fact be distributed and/or hierarchical, as is known in the art. For example, the storage device 115 may exist in multiple levels of caches, and these caches may be further divided by function, so that one cache holds instructions while another holds non-instruction data which is used by the processor or processors. The storage device 115

may further be distributed and associated with different processors or sets of processors, as is known in any of various so-called non-uniform memory access (NUMA) computer architectures. Further, although the electronic device 102 is drawn to contain the storage device 115, it may be distributed across other electronic devices, such as devices
5 connected to the network 108.

The storage device 115 includes a data assembler 126, a data normalizer 127, a data dictionary 128, a model norm 130, a supplier registry 132, records 134, and a report 136, all of which may in various embodiments have any number of instances. Although the data assembler 126, the data normalizer 127, the data dictionary 128, the model norm
10 130, the supplier registry 132, the records 134, and the report 136 are all illustrated as being contained within the storage device 115 in the electronic device 102, in other embodiments some or all of them may be on different electronic devices and may be accessed remotely, e.g., via the network 108.

The data assembler 126 registers the information suppliers 104 in the supplier
15 registry 132 and creates the model norm 130. In an embodiment, the data assembler 126 selects the information suppliers 104 based on the opinions of the designers of the data assembler 126, based on input from users, based on content of a publicly accessible registry, or based on any combination thereof. The data assembler 126 receives requests for information from the client 106 and in response finds the appropriate information
20 suppliers 104 using the supplier registry 132. The data assembler 126 then sends a request to the appropriate information supplier(s) 104 and receives records from the information supplier(s) 104, which the data assembler 126 stores as the records 134. The data normalizer 127 then creates the report 136 based on the records 134 using the model norm 130 and sends the report 136 to the client 106 in response to the client's request.

25 In an embodiment, the data assembler 126 and the data normalizer 127 include instructions capable of executing on the processor 110 or statements capable of being interpreted by instructions executing on the processor 110 to carry out the functions as further described below with reference to Fig. 7. In another embodiment, the data assembler 126 and/or the data normalizer 127 may be implemented in hardware via logic

gates and/or other appropriate hardware techniques in lieu of or in addition to a processor-based system.

The supplier registry 132 is further described below with reference to Fig. 2. The data dictionary 128 is further described below with reference to Fig. 3. The model norm
5 130 is further described below with reference to Fig. 5. The report 136 is further described below with reference to Fig. 6B.

The input device 120 may be a keyboard, mouse or other pointing device, trackball, touchpad, touchscreen, keypad, microphone, voice recognition device, or any other appropriate mechanism for the user to input data to the electronic device 102 and/or
10 to manipulate the user interfaces of the electronic device 102. Although only one input device 120 is shown, in another embodiment any number and type of input devices may be present. The input device 120 may be used to interact with and manipulate the user interfaces of the electronic device 102, if any.

The output device 122 is that part of the electronic device 102 that presents output
15 to the user. The output device 122 may be a cathode-ray tube (CRT) based video display well known in the art of computer hardware. But, in other embodiments the output device 122 may be replaced with a liquid crystal display (LCD) based or gas, plasma-based, flat-panel display. In still other embodiments, any appropriate display device may be used. In other embodiments, a speaker or a printer may be used. In other
20 embodiments any appropriate output device may be used. Although only one output device 122 is shown, in other embodiments, any number of output devices of different types or of the same type may be present. The output device 122 may display or otherwise present the user interfaces of the electronic device 102, if any.

The bus 125 may represent one or more busses, e.g., PCI (Peripheral Component
25 Interconnect), ISA (Industry Standard Architecture), X-Bus, EISA (Extended Industry Standard Architecture), or any other appropriate bus and/or bridge (also called a bus controller). Although the bus 125 is shown in Fig. 1 as a relatively simple, single bus structure providing a direct communication path between the processor 110, the storage device 115, the input device 120, and the output device 122, in other embodiments the

bus 125 may comprise multiple different buses or communication paths, which may be arranged in any of various forms, such as point-to-point links in hierarchical, star or web configurations, multiple hierarchical buses, or parallel and redundant paths. Furthermore, while the bus 125 is shown directly connected to the processor 110, the storage device 115, the input device 120, and the output device 122, in other embodiments, some or all of the I/O (Input/Output) devices may be connected via I/O processors.

The electronic device 102 may be implemented using any suitable hardware and/or software, such as a personal computer. Portable computers, laptop or notebook computers, PDAs (Personal Digital Assistants), pocket computers, telephones, pagers, automobiles, teleconferencing systems, appliances, and mainframe computers are examples of other possible configurations of the electronic device 102. The hardware and software depicted in Fig. 1 may vary for specific applications and may include more or fewer elements than those depicted. For example, other peripheral devices such as audio adapters, or chip programming devices, such as EPROM (Erasable Programmable Read-Only Memory) programming devices may be used in addition to or in place of the hardware already depicted.

The information supplier 104 may be an electronic device including hardware and optional software components analogous to the electronic device 102 previously described above. The information supplier 104 may function as a server and includes a favored norm 160 and records 162. The information supplier 104 registers its area of interest with the electronic device 102 and supplies the favored norm 160 and records 162 in response to a request from the electronic device 102. Although the favored norm 160 and the records 162 are illustrated as being separate entities, in another embodiment they may be combined. The favored norm 160 indicates the information supplier's 104 view of how the characteristics in the records 162 should be organized and how the records 162 should be sorted. The favored norm 160 is further described below with reference to Fig. 4. A data structure for the records 162 is further described below with reference to Fig. 6A.

The client 106 may be an electronic device including hardware and optional software components analogous to the electronic device 102 previously described above.

The client 106 sends requests for information to the electronic device 102 and receives reports from the electronic device 102.

The network 108 may be any suitable network or combination of networks and may support any appropriate protocol suitable for communication of data and/or code
5 between the electronic device 102, the information supplier 104, and the client 106. In various embodiments, the network 108 may represent a storage device or a combination of storage devices. In an embodiment, the network 108 may support Infiniband. In another embodiment, the network 108 may support wireless communications. In another embodiment, the network 108 may support hard-wired communications, such as a
10 telephone line or cable. In another embodiment, the network 108 may support the Ethernet IEEE (Institute of Electrical and Electronics Engineers) 802.3x specification. In another embodiment, the network 108 may be the Internet and may support IP (Internet Protocol). In another embodiment, the network 108 may be a local area network (LAN) or a wide area network (WAN). In another embodiment, the network 108 may be a
15 hotspot service provider network. In another embodiment, the network 108 may be an intranet. In another embodiment, the network 108 may be a GPRS (General Packet Radio Service) network. In another embodiment, the network 108 may be a FRS (Family Radio Service) network. In another embodiment, the network 108 may be any appropriate cellular data network or cell-based radio network technology. In another
20 embodiment, the network 108 may be an IEEE 802.11B wireless network. In still another embodiment, the network 108 may be any suitable network or combination of networks. Although one network 108 is shown, in other embodiments any number of networks (of the same or different types) may be present.

The various software components illustrated in Fig. 1 and implementing various
25 embodiments of the invention may be implemented in a number of manners, including using various computer software applications, routines, components, programs, objects, modules, data structures, etc., referred to hereinafter as "computer programs," or simply "programs." The computer programs typically comprise one or more instructions that are resident at various times in various memory and storage devices in the electronic device
30 102, and that, when read and executed by one or more processors in the electronic device

102, cause the electronic device 102 to perform the steps necessary to execute steps or elements embodying the various aspects of an embodiment of the invention.

Moreover, while embodiments of the invention have and hereinafter will be described in the context of fully functioning electronic devices, the various embodiments
5 of the invention are capable of being distributed as a program product in a variety of forms, and the invention applies equally regardless of the particular type of signal-bearing medium used to actually carry out the distribution. The programs defining the functions of this embodiment may be delivered to the electronic device 102 via a variety of signal-bearing media, which include, but are not limited to:

10 (1) information permanently stored on a non-rewriteable storage medium, e.g., a read-only memory device attached to or within an electronic device, such as a CD-ROM readable by a CD-ROM drive;

(2) alterable information stored on a rewriteable storage medium, e.g., a hard disk drive or diskette; or

15 (3) information conveyed to an electronic device by a communications medium, such as through a computer or a telephone network, e.g., the network 108, including wireless communications.

Such signal-bearing media, when carrying machine-readable instructions that direct the functions of the present invention, represent embodiments of the present
20 invention.

In addition, various programs described hereinafter may be identified based upon the application for which they are implemented in a specific embodiment of the invention. But, any particular program nomenclature that follows is used merely for convenience, and thus embodiments of the invention should not be limited to use solely
25 in any specific application identified and/or implied by such nomenclature.

The exemplary environments illustrated in Fig. 1 are not intended to limit the present invention. Indeed, other alternative hardware and/or software environments may be used without departing from the scope of the invention.

Fig. 2 depicts a block diagram of an example data structure for the supplier registry 132, according to an embodiment of the invention. The supplier registry 132 includes entries 205, 210, and 215, but in other embodiments any number of entries may be present. Each entry includes a supplier 220 and an area of interest 225. The supplier
5 220 indicates a supplier of information, such as the information supplier 104. The area of interest 225 indicates an area or category of information that the supplier 220 is capable of providing. Examples of areas of interest as illustrated in Fig. 2 are cars, fishing, and market research, but in other embodiments any appropriate areas of interest may be used.

Fig. 3 depicts a block diagram of an example data structure for the data dictionary
10 128, according to an embodiment of the invention. In an embodiment, one data dictionary 128 includes information regarding all areas of interest, but in another embodiment each area of interest has its own associated data dictionary 128. The data dictionary 128 describes a data definition for fields in a record of information, such as the records 134. The data dictionary 128 includes entries 302, 304, 306, and 308, but in
15 other embodiments any number of entries may be present. Each entry includes a characteristic 310 and a default 312. The characteristic 310 identifies a field in the records 134. In the example shown in Fig. 3, the fields in the records 134 are a name field, a zip code field, a state field, and an income field, but in another embodiments any appropriate fields may be identified in the characteristic 310. The default 312 indicates a
20 default value for the field identified by the associated characteristic 310. In the example shown in Fig. 3, the name field has a default value of "John Doe," the zip code field has a default value of "55901," the state field has a default value of "Minnesota," and the income field has a default of "\$0", but in other embodiments any appropriate default values may be used. In other embodiments, the entries in the data dictionary 128 may
25 include more elements in addition to the characteristic 310 and the default 312.

In an embodiment, the format and contents of data dictionary 128 may be standardized by an agreement among the information suppliers 104. The standardization of the characteristics 310 and their data-types and usage allows the data normalizer 127 to create the report 136 by matching the characteristics in the model norm 130 to the

characteristics in the favored norm 160, as further described below with reference to Figs. 4, 5, and 7.

Fig. 4 depicts a block diagram of an example data structure for the favored norm 160, according to an embodiment of the invention. The favored norm 160 includes entries 405, 410, 415, and 420, but in other embodiments any appropriate number of records may be present. Each entry includes a characteristic 422, a relative significance 424, a default value 426, and a sort rule 428. The characteristic 422 identifies a field in the records 134. The relative significance 424 indicates the information supplier's 104 opinion or view of the relative significance of the associated characteristic 422, i.e., the relative order of the associated characteristic 422 with respect to other characteristics. The default value 426 indicates the information supplier's 104 opinion or view of the default value of the associated characteristic 422. In other embodiments, the default value 426 is optional or not present. The data normalizer 127 may generate a default value, even for characteristics not drawn from the data dictionary 128, based on the characteristic's data type. The sort rule 428 indicates the information supplier's 104 opinion or view of how the records 134 are to be sorted. In an embodiment, the sort rule 428 indicates the information supplier's 104 view of the sort order of the records 162.

Fig. 5 depicts a block diagram of an example data structure for the model norm 130, according to an embodiment of the invention. The model norm 130 includes entries 505, 510, and 515 but in other embodiments any number of entries may be present. Each entry includes a characteristic 522, a relative significance 524, a default value 526, and a sort rule 528. The characteristic 522 identifies an associated field in the records 134. The data normalizer 127 uses the characteristic 522 to determine which fields to select from the records 134 and include in the report 136. The relative significance 524 indicates a relative significance of the associated characteristic 522, and the data normalizer 127 uses the relative significance 524 to order the characteristics in the report 136 as further described below with reference to Fig. 7. The default value 526 indicates a default of the associated characteristic 522, which the data normalizer 130 uses to fill in the report 136 if the associated characteristic does not have an assigned value in the records 134, as further described below with reference to Fig. 7. The sort rule 528

indicates how records are to be sorted, and the data normalizer 127 uses the sort rule 528 to sort the records in the report 136, as further described below with reference to Fig. 7. In an embodiment, the sort rule 528 may indicate that records are to be sorted in ascending order, descending order, or in any other appropriate order. The sort rule 528 may also indicate whether the associated characteristic is to be used as a sort field and any associated sorting information, e.g., the direction of the sort such as ascending or descending and a sort priority.

Fig. 6A depicts a block diagram of an example data structure for the records 162 prior to being normalized by the data normalizer 127, according to an embodiment of the invention. The records 162 include records 605, 610, 615, and 620, but in other embodiments any number of records with any appropriate data may be present. Each record includes characteristics 621, 622, 624, and 626, but in other embodiments any number of characteristics with any appropriate characteristic names may be used.

Fig. 6B depicts a block diagram of an example data structure for the report 136 after being normalized by the data normalizer 127, according to an embodiment of the invention. The report 136 includes records 655, 660, 665, and 670, but in other embodiments any number of records may be present. In an embodiment, all of the records 655, 660, 665, and 670 originate from the same information supplier 104. In another embodiment, some or all of the records 655, 660, 665, and 670 originate from different information suppliers 104.

Each record includes characteristics 672, 674, and 676, which the data normalizer 127 selects from the records 134 because they exist in the model norm 130. The data normalizer 127 orders the characteristics 672, 674, and 676 in the report 136 based on the relative significance 524 in the model norm 130, which may have a different order from the records 162, e.g., as shown in Fig. 6A. Thus, using the example of the model norm 130 in Fig. 5, the zip code characteristic 672 has a relative significance 524 in the model norm 130 of "1," so the zip code characteristic 672 is ordered first (placed in the leftmost column) in the report 136. The income characteristic 674 has a relative significance 524 in the model norm 130 of "2," so the income characteristic 674 is ordered second (placed in the centermost column) in the report 136. The state characteristic 676 has a relative

significance 524 of “3,” in the model norm 130, so the state characteristic 670 is ordered third (placed in the rightmost column) in the report 136. In other embodiments, the characteristics may be ordered right-to-left, up-to-down, down-to-up, or in any other appropriate order in the report 136 based on the relative significance 524 in the model norm 130.

The data normalizer 127 has not included the name characteristic 621 (Fig. 6A) in the report 136, because it does not exist in the model norm 130. Thus, the name characteristic 621 is an extra characteristic because it is included in the favored norm 160 (as shown in Fig. 4) but not in the model norm 130 (as shown in Fig. 5). In the embodiment shown in Fig. 6B, the data normalizer 127 has omitted the extra characteristic (in this example the name characteristic 621).

In another embodiment, the data normalizer 127 presents each record’s extra characteristic(s) as a sequence of name-value pairs in a single least-significant column ordered in the sequence within the column as indicated by the relative significance 424 in the favored norm 160 (Fig. 4). This embodiment may be especially useful when the report 136 is made up of information from multiple information suppliers 104.

In another embodiment, the data normalizer 127 presents any extra characteristic(s) in a series of less-significant columns ordered as indicated by the relative significance 424 in the favored norm 160 (Fig. 4). This embodiment may be especially useful when the report 136 is made up of information from a single information supplier 104.

In another embodiment, the data normalizer 127 provides a single least-significant column in the report 136 containing a visual indicator of whether extra characteristics are available for each report record.

The data normalizer 127 orders the records 655, 660, 665, and 670 based on the sort rule 528 in the model norm 130, which may have a different order from the records 162, e.g., as illustrated in Fig. 6A. Using the example of Fig. 5, the sort rule 528 for the income characteristic 674 is “desc.,” which indicates that the income characteristic is to be used as a sort field and the records are to be in descending order. Thus, the data

normalizer 127 orders the records 655, 660, 665, and 670 within the report 136 from the highest income at the top to the lowest income at the bottom, but in other embodiments any appropriate sort order of the records may be used.

Fig. 7 depicts a flowchart of example processing, according to an embodiment of the invention. Control begins at block 700. Control then continues to block 705 where the information supplier 104 registers with the data assembler 126. The data assembler 126 creates an entry in the supplier registry 132 for each registered information supplier 104 and stores an identifier of the information supplier 104 in the supplier 220 in the entry and an identifier of the area of interest associated with the information supplier 104 in the area of interest 225 in the entry. The data assembler 126 also creates the data dictionary 128 based on common characteristics for the field of interest among the information suppliers 104.

Control then continues to block 710 where the client 106 requests a report on a field of interest from the electronic device 102.

Control then continues to block 715 where the data assembler 126 consults the supplier registry 132 and finds a match in the area of interest 225 for the client's area of interest and requests records from the associated information supplier 220 in the found entry, such as the information supplier 104. Control then continues to block 720 where the information supplier 104 returns the records 162 and the favored norm 160 to the data assembler 126.

Control then continues to block 725 where the data normalizer 127 creates the model norm 130. In various embodiments, the data normalizer 127 may create the model norm 130 based on the favored norm 160, based on the data dictionary 128, based on the judgment of the designer of the data normalizer 127, based on any other appropriate criteria, or based on any combination thereof.

Control then continues to block 730 where the data normalizer 127 selects characteristics from the records 134 based on the characteristics specified in the model norm 130 and orders the selected characteristics in the report 136 based on the relative significance 524 in the model norm 130. Control then continues to block 735 where the

data normalizer 127 orders the records 134 in the report 136 based on the sort rule 528 in the model norm 130. The data normalizer 127 further propagates the records with the values in the default values 526 if the records lack values. Control then continues to block 740 where the data normalizer 127 sends the report 136 to the client 106.

5 Control then continues to block 745 where the client 106 selects a record in the report 136 and requests a new report. Control then continues to block 750 where the data normalizer 127 creates a new report based on the favored norm 160 associated with the selected record. The data normalizer 127 selects the characteristics specified in the favored norm 160 and orders them in the new report based on the relative significance
10 424 in the favored norm 160 associated with the selected record. The data normalizer 127 further sorts the records in the new report based on the sort rule 428 in the favored norm 160 associated with the selected record. In another embodiment, the data normalizer 127 creates the new report based on a preferred norm supplied by the client 106. Control then continues to block 799 where the logic returns.

15 In the previous detailed description of exemplary embodiments of the invention, reference was made to the accompanying drawings (where like numbers represent like elements), which form a part hereof, and in which is shown by way of illustration specific exemplary embodiments in which the invention may be practiced. These embodiments were described in sufficient detail to enable those skilled in the art to practice the
20 invention, but other embodiments may be utilized and logical, mechanical, electrical, and other changes may be made without departing from the scope of the present invention. Different instances of the word “embodiment” as used within this specification do not necessarily refer to the same embodiment, but they may. The previous detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present
25 invention is defined only by the appended claims.

 In the previous description, numerous specific details were set forth to provide a thorough understanding of the invention. But, the invention may be practiced without these specific details. In other instances, well-known circuits, structures, and techniques have not been shown in detail in order not to obscure the invention.